Gene cooption without duplication during the evolution of a male-pregnancy gene in pipefish

April Harlin-Cognato[†‡], Eric A. Hoffman[§], and Adam G. Jones[†]

[†]Department of Biology, Texas A&M University, 3258 TAMU, College Station, TX 77843-3258; and [§]Department of Biology, University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816

Edited by Morris Goodman, Wayne State University School of Medicine, Detroit, MI, and approved October 5, 2006 (received for review April 14, 2006)

Comparative studies of developmental processes suggest that novel traits usually evolve through the cooption of preexisting genes and proteins, mainly via gene duplication and functional specialization of paralogs. However, an alternative hypothesis is that novel protein function can evolve without gene duplication, through changes in the spatiotemporal patterns of gene expression (e.g., via cis-regulatory elements), or functional modifications (e.g., addition of functional domains) of the proteins they encode, or both. Here we present an astacin metalloprotease, dubbed patristacin, which has been coopted without duplication, via alteration in the expression of a preexisting gene from the kidney and liver of bony fishes, for a novel role in the brood pouch of pregnant male pipefish. We examined the molecular evolution of patristacin and found conservation of astacin-specific motifs but also several positively selected amino acids that may represent functional modifications for male pregnancy. Overall, our results pinpoint a clear case in which gene cooption occurred without gene duplication during the genesis of an evolutionarily significant novel structure, the male brood pouch. These findings contribute to a growing understanding of morphological innovation, a critically important but poorly understood process in evolutionary biology.

novel trait evolution | patristacin | Syngnathidae

Evolutionary innovation has been defined as "the origin of a novel body part which may serve a novel function or specialize in a function that was already performed in the ancestral lineage but without a dedicated organ" (ref. 1, p. 581). In seahorses and pipefishes (family Syngnathidae) males carry their embryos on their ventral surface, either exposed to the environment or enclosed in a fleshy brood pouch (Fig. 1). The brood pouch is a clear example of an evolutionary innovation: syngnathids are the only lineage to have evolved a morphological structure that allows males to become pregnant. In many species of syngnathid fishes, the brood pouch is a complex organ composed of highly vascularized epithelial tissue that forms a honeycomb matrix encapsulating individual eggs during gestation (Fig. 1) (2). This placenta-like tissue apparently supplies nutrients to developing embryos (3) in a manner analogous to the placenta of female mammals. The pouch is lined with cells rich in mitochondria (CRMs) (Fig. 1) (2, 4) that transfer ions between the brood pouch fluid and the male bloodstream, maintaining an osmotically neutral environment during the early stages of gestation (4–6). A phylogeny of the Syngnathidae suggests two independent origins of the brood pouch based on its location on either the abdomen (Gastrophori) or the tail (Urophori) of the male (7), and within these two lineages there is considerable variation in the degree to which the pouch encloses developing embryos, the complexity of the "pseudoplacenta," and the structure of brood pouch folds (7).

Clearly, the evolution of a structure as complex as a brood pouch required the coevolution of a multitude of genes, but little is known about the genomic changes associated with the evolution of the developmental and physiological processes that allow such a unique feature to function. The conservation of proteins and motifs across the tree of life suggests that the appropriation, or cooption, of preexisting genes and proteins plays an important role in the evolution of novel traits (8). Gene cooption occurs when a new function is derived for an old gene, usually through changes in expression patterns and/or functional modifications of the protein it encodes (8). Because protein evolution is necessarily constrained to retain ancestral function, it has long been considered that duplications of genes, gene segments, or entire genomes are the primary source of material for the cooption of proteins during the evolution of novel traits (9, 10). However, it is not clear how duplicated genes survive the accumulation of point mutations while making the transition to a new function (10) or how these new functions are acquired (11). A number of studies have shown that duplicated genes can

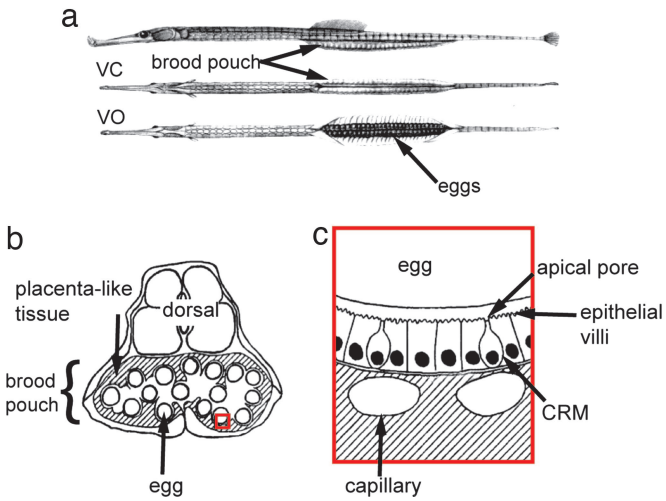

Fig. 1. Diagram of brood pouch morphology of tail-brooding pipefish. (*a*) Three views of the brood pouch. VC, ventral, closed; VO, ventral, opened. (*b*) A cross-section of a gravid brood pouch, showing the placenta-like epithelium with capsules for individual embryos. Note the complete closure of the pouch folds. (*c*) A closer view shows details of the structure of the pseudo placenta and the egg–placenta interface. The egg lies on the surface of villous epithelial cells interspersed with CRMs that have an apical pore opening into the lumen and a base that lies close to capillaries. These CRMs are thought to be the conduits by which ions are transported out of the brood pouch to maintain an environment suitable for developing embryos. This figure is adapted from Kornienko (41).

Author contributions: A.H.-C., E.A.H., and A.G.J. designed research; A.H.-C. and E.A.H. performed research; A.H.-C. analyzed data; and A.H.-C. and A.G.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: CRMs, cells rich in mitochondria; BF, Bayes factor; REL, random effects likelihood.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF060269–EF060286).

[‡]To whom correspondence should be sent at the present address: Department of Zoology, Michigan State University, 203 Natural Sciences Building, East Lansing, MI 48824. E-mail: cognatoa@msu.edu.

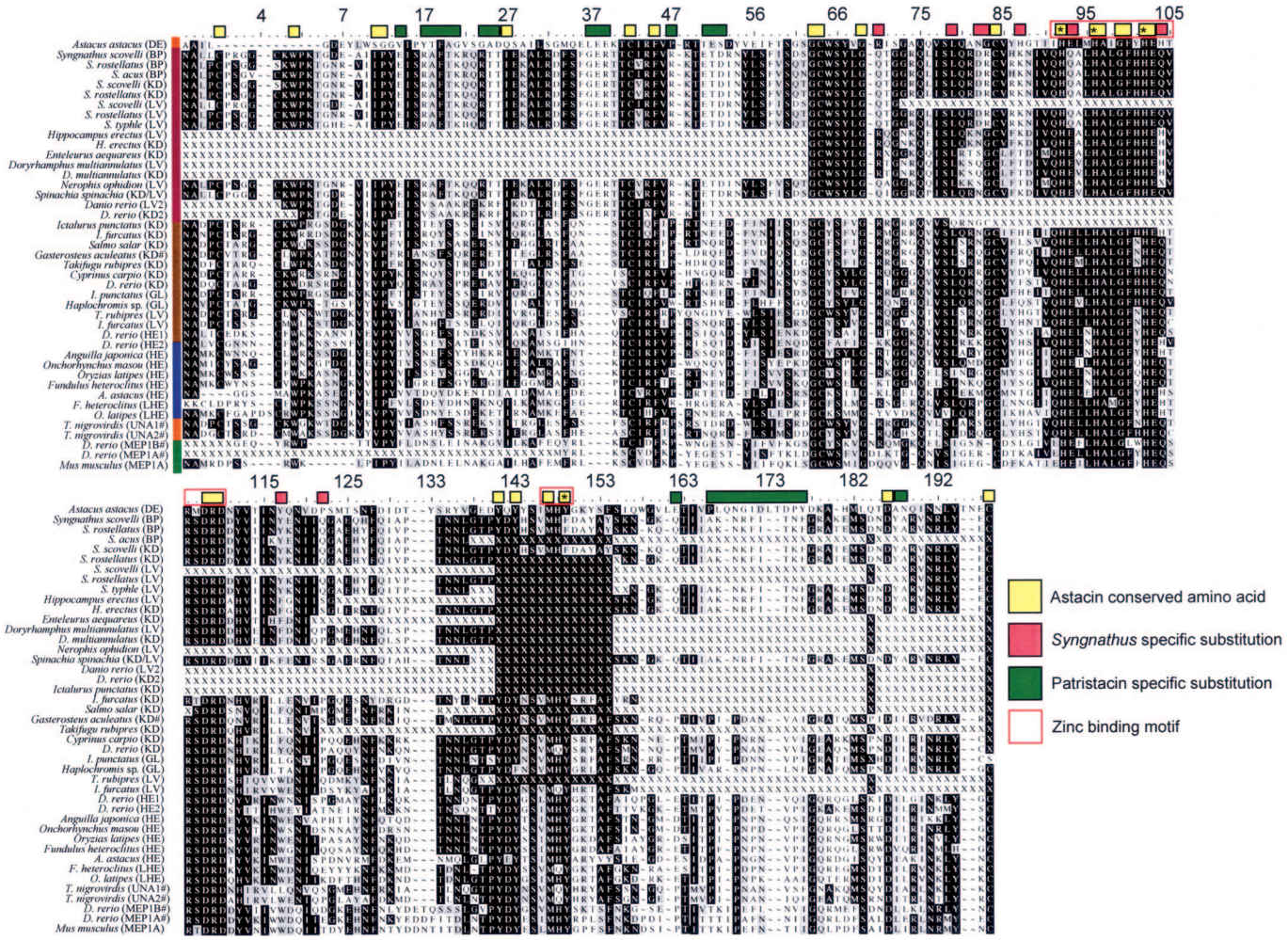

Fig. 2. Amino acid alignment of astacin mature peptides. Amino acids are numbered according to alignment with astacin, the type-protein of the subfamily. Shaded blocks designate amino acid positions that share ≥35% similarity across all sequences. Astacins are characterized by four zinc bonds (indicated by asterisks) within two conserved zinc binding motifs (HEXXHAXGFXGEXRXDR, MHY) (17, 42). Colored bars to the left of the sequences identify astacin protein type as in Fig. 3. DE, digestive enzyme; BP, brood pouch; LV, liver; KD, kidney; GL, gill; HE, high hatching enzyme; LHE, low hatching enzyme; UNA, unknown tissue type; MEP, meprin subunits α (A) and β (B); #, proteins whose tissue-specific association was inferred from phylogenetic analysis; X, missing data; ~, gap inserted for alignment.

survive in the genome by acquiring a new function (neofunctionalization) or by partitioning their function among gene copies (subfunctionalization) (10, 12, 13). Some examples of these processes include the addition of functional domains to an existing protein, or changes in cis-regulatory elements that alter the spatiotemporal patterns of protein expression (13). However, only a fraction of gene duplications are thought to make the transition to new functions (14), so it is unlikely that biological diversity can be explained entirely by the processes of duplicate gene evolution.

As an alternative to the strict gene duplication model, gene cooption may occur through the alteration of spatial patterns of expression of existing genes. Under this model, functional promiscuity and conformational diversity of proteins play an important role in the evolution of novelty (11, 15). Proteins are evolutionarily labile, and variation in protein conformation and substrate affinity in different physiochemical environments and in the presence of a variety of substrates allows a preexisting protein to rapidly acquire a functional role associated with a novel structure (11, 15, 16), without gene duplication or changes to the original protein function.

Our goals in the present study were to identify genes involved in male pregnancy and to investigate the molecular evolution of one of these male-pregnancy genes to distinguish between different models of gene cooption. In particular, we wanted to test the hypotheses that gene cooption during the evolution of male pregnancy was accompanied by gene duplication, and that genes involved in male pregnancy experienced a history of positive selection associated with their recent cooption to the brood pouch. Under the duplication model, we expect to recover a brood-pouch-specific gene lineage immediately after a duplication event, whereas under the nonduplication model we expect there to be no indication of tissue-specific gene evolution. Positive selection is possible under both duplication and nonduplication models and would indicate potential modifications to the protein associated with male pregnancy. However, the interpretation of these results would differ under each model.

Results and Discussion

To investigate the genetic basis of novel gene function in the brood pouch of pipefishes and seahorses, we used subtractive suppression hybridization of cDNA libraries from the brood pouch tissue of pregnant and nonpregnant male Gulf pipefish (*Syngnathus scovelli*). This approach allowed us to construct a cDNA library enriched for genes that were up-regulated in the

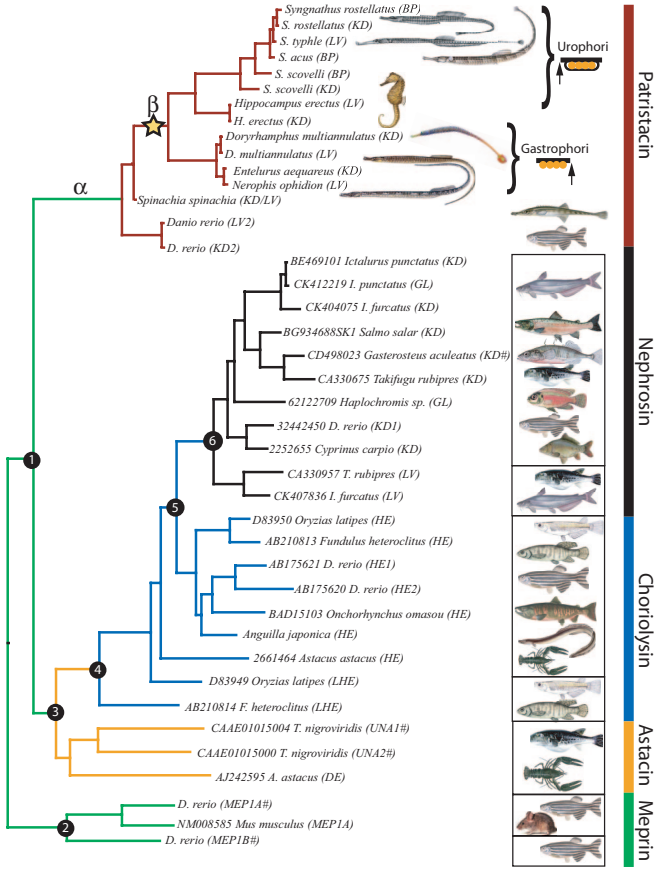

Fig. 3. Evolutionary history of patristacin and other astacin proteins. Hypothesized point of cooption of patristacin for male pregnancy is indicated by a star. Numbers 1–6 indicate inferred gene duplication events derived from shared ancestry of tissue-specific expression. Branches are colored according to protein type. Tests for change in selection pressures were performed for lineages after the duplication event from which patristacin originated (α) and was coopted for male pregnancy (β). Within the Syngnathidae, the Urophori and Gastrophori are monophyletic clades defined by position of the pouch either anterior or posterior to the anal vent. The Gastrophori represented in this study had the simplest of brooding structures; eggs are glued to the ventral surface of the male without enclosure in a pouch. GenBank accession numbers are provided for sequences used in this study (Table 1), and tissue types from which sequences were collected are denoted as in Fig. 2. #, proteins whose tissue-specific association was unknown before phylogenetic analysis.

pregnant male's brood pouch. The most highly represented, differentially expressed transcript in our cDNA library was a novel gene product, which we call "patristacin." We used real-time quantitative RT-PCR to verify that patristacin is indeed up-regulated during male pregnancy (Kruskal–Wallis, $P = 0.007$) (Fig. 5, which is published as supporting information on the PNAS web site). Patristacin contains the two conserved zinc binding motifs (HEXXHAXGFXGEXRXDR, MHY) that are diagnostic of astacins, a subfamily of zinc-dependent metalloproteases (17). Recent studies have characterized multiple tissue-specific astacin proteins in a variety of organisms from crayfish to mammals. The astacin gene family includes digestive enzymes (18), hatching enzymes (19, 20), bone morphogenic proteins (21), meprins (22), and the recently discovered proteins cimp1 (23) and nephrosin (24), which are expressed in the gills and kidney of bony fishes. We obtained the complete coding sequence of patristacin from the brood pouches of pregnant male *S. scovelli*, *Syngnathus rostellatus*, and *Syngnathus acus*, as well as representative sequences from the kidney and liver of several syngnathids and other fishes (Table 1, which is published as supporting information on the PNAS web site). An alignment of amino acid sequences revealed that the substrate binding and structural residues of patristacin and other astacins are remarkably conserved across the protein subfamily (Fig. 2). Outside of these conserved motifs, however, there are several amino acid substitutions unique to patristacin (Fig. 2), including a diagnostic 3-aa insertion near the 5′ end of the protein, which is shared among all patristacin coding sequences regardless of species and tissue type. The presence of patristacin transcripts in the liver and kidney of multiple fish species, including seahorses and pipefishes, suggests that this protein has an ancient origin in bony fishes, perhaps with a role in liver and kidney function.

To investigate the evolutionary origin of patristacin and its role in the brood pouch of pregnant male pipefishes, we reconstructed a phylogeny of patristacin and other astacin proteins from a wide selection of taxa (Fig. 3). The resulting phylogeny suggests that there have been at least six duplications of astacin proteins in the teleosts, each followed by tissue-specific specialization of paralogs (Fig. 3). For example, choriolysin proteins have two isoforms ("high" and "low") that originated through gene duplication and now function together to break down the egg chorion before hatching (19, 20). Likewise, our results suggest that nephrosin, a relatively new member of the astacin subfamily, has two forms (i.e., gill/kidney and liver) that were each coopted for tissue-specific expression after gene duplication (Fig. 3). Hence, multiple astacins, including patristacin, which share a single common ancestor, are expressed in various tissues in a single individual at any given time.

Several lines of evidence suggest that patristacin was expressed in at least the kidney and liver before the origin of male pregnancy and that it expanded its spatial pattern of expression, without gene duplication, to include the brood pouch once male pregnancy evolved. First, the observation that patristacin transcripts are present in the liver and kidney of multiple fish species without male pregnancy, including the fifteenspine stickleback (*Spinachia spinachia*) and zebrafish (*Danio rerio*), suggests that this protein had an ancient origin in bony fishes, perhaps with a role in liver and kidney function. Second, patristacin in the Syngnathidae does not form monophyletic lineages based on tissue-specific expression (Fig. 3). If cooption of patristacin to the brood pouch had been accompanied by gene duplication, then we would have expected to recover a monophyletic, brood pouch-specific gene clade and a related, monophyletic kidney and liver clade. Instead, we find that the liver and kidney sequences are intermingled with the brood pouch sequences in the gene phylogeny within the genus *Syngnathus* (Fig. 3). In those cases where more than one patristacin per species is represented, the kidney, liver, and brood pouch sequences from a given species are more similar to one another than they are to sequences from any tissue type of a related species. For example, in *S. rostellatus*, the kidney sequence displays only one amino acid substitution relative to the brood pouch sequence. We found a similar pattern in *Syngnathus typhle* (data not shown). One interesting pattern, however, is that we observed more divergence among patristacin sequences from different tissues within *S. scovelli* than we observed among tissues in other species in the genus. The divergence could be explained by a large amount of allelic variation at patristacin within *S. scovelli*. Alternatively, there may have been a recent gene duplication on the lineage leading to *S. scovelli*, after the radiation of the genus *Syngnathus*. More data will be needed to resolve this issue, but our data nevertheless are inconsistent with the hypothesis that a gene duplication coincided with the patristacin cooption event, an occurrence that clearly predated the diversification of the genus *Syngnathus*. In summary, our data support the hypothesis that the cooption of patristacin for use by the brood pouch

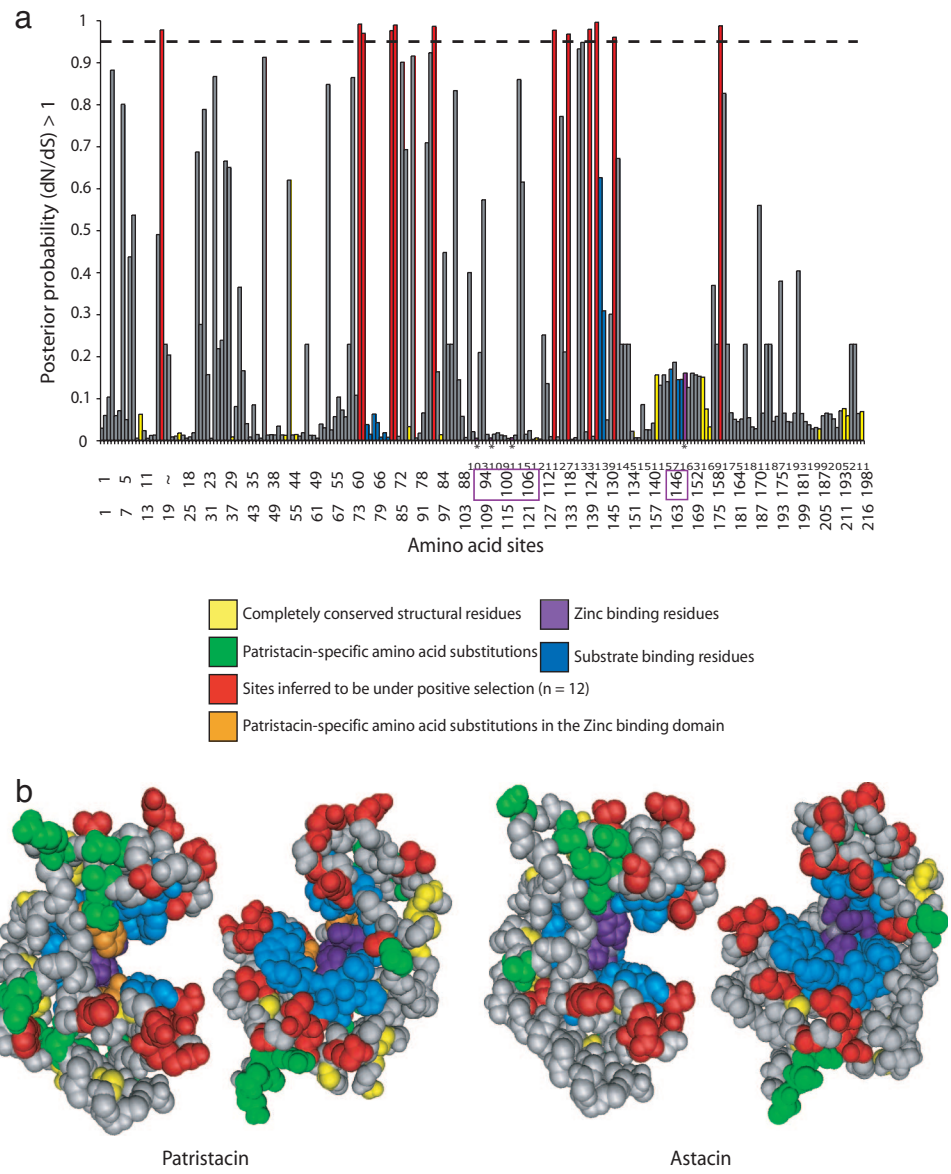

Fig. 4. Comparative modeling of patristacin amino acid substitutions and sites under positive selection. (*a*) Graphical representation of selection pressures among patristacin amino acid sites. Amino acids in patristacin are numbered 1–216 (bottom numbers) and with corresponding amino acids of astacin (top numbers), the type protein for the astacin subfamily. *, the position of zinc binding residues. Boxed amino acid numbers designate motifs that are diagnostic of astacins. Sites that met both posterior probability (PAML analysis; above dotted line) and/or high BF (REL analysis) criteria (see *Materials and Methods*) were considered to be under positive selection ($n = 12$). (*b*) 3D models of patristacin and astacin in two orientations. Amino acids on the astacin model are colored for comparison to sites in the patristacin model. Note the clustering of sites under positive selection (highlighted red) around the functional cleft of the protein.

occurred during the evolution of male pregnancy without gene duplication or loss of the protein's ancestral function. Precise dating of the cooption event relative to the evolution of male pregnancy will require additional data on the patristacin gene phylogeny with respect to the Syngnathidae species tree.

One interpretation of these results is that patristacin performs a function in the brood pouch similar to that in the kidney and liver but acquired a novel role in male pregnancy via a change in the spatiotemporal patterns of gene expression. If this hypothesis is correct, we expect very little modification to amino acid domains that are important to the ancestral structure and function of the protein. To test this hypothesis, we used maximum-likelihood-based analyses to examine whether the patterns of molecular evolution of patristacin after cooption are consistent with functional or structural changes in the protein. Because we were interested in changes in patristacin after cooption for male pregnancy, we compared patristacin sequences from the Syngnathidae to a subset of sequences from other astacins. If there had been an acceleration in protein evolution associated with gene cooption, then we expect a model that allows variation in rates of evolution to explain our data better than a model that does not. We tested two models that allowed accelerated rates of evolution either at the base of the patristacin clade (α in Fig. 3) or on the branch leading to the Syngnathidae (β in Fig. 3). Neither of the two ratio models were significantly different from the one ratio model (Table 2, which is published as supporting information on the PNAS web site), which suggests that the rate of evolution along lineages leading to the duplication and cooption of patristacin (lineages α and β) (Fig. 3) is not different from the background substitution rate of astacins. These results support a relatively constant rate of evolution across all astacin gene lineages (Table 2).

Our examination of site-specific selection pressures revealed a general pattern of strong purifying selection in the functional regions of the astacin proteins, including patristacin (Fig. 4). In addition, these tests identified 12 amino acid sites in patristacin that are under positive selection (see *Materials and Methods*), 84% of which are within 30 aa of functional motifs (Fig. 4). A 3D model of patristacin, derived from comparison of the complete coding sequence of *S. scovelli* brood pouch patristacin to the fully resolved structural model of astacin (see *Materials and Methods*), indicates that the sites under positive selection are clustered around the outer rim of the functional site (Fig. 4). Previous studies suggest that amino acid changes adjacent to substrate binding motifs in meprin α and astacin generally do not alter the substrate specificity of the proteins but could affect the rate at which the active site interacts with substrates (17). In addition, some amino acid substitutions have resulted in small, but potentially important, modifications to the tertiary structure of the protein (Fig. 4). For example, when compared with astacin, several sites in patristacin have been substituted with amino acids that have nonequivalent physiochemical properties (25), which may have altered the conformation of the protein in some cases (Fig. 4). Whether these changes have altered the function of patristacin is unclear, but our results do suggest that, at least in the Syngnathidae, patristacin has been shaped to some extent by positive selection, perhaps in a manner that has altered the efficiency or substrate binding affinity of the protein. These changes in protein efficiency could be a direct result of selection pressures associated with the cooption of patristacin for male pregnancy, but more extensive sampling of complete patristacin sequences from other taxa are required before we can draw definitive conclusions regarding the evolution of patristacin and male pregnancy.

Although the actual function of patristacin in the brood pouch is unknown, there are several possibilities that can be derived from what we know about the function of other astacins in fishes. For example, fishes osmoregulate via their gills and kidneys, which contain an abundant supply of CRMs (26, 27), similar to those that regulate ion concentration in the brood pouch during gestation. Therefore, it is possible that meprins, cimp1, nephrosin, and patristacin, a group of closely related metalloproteases expressed in organs with CRMs, all perform a function in osmoregulation. Another intriguing possibility is that patristacin may have evolved a function similar to the related hatching enzymes. In oviparous species, choriolytic enzymes are produced by the embryo to facilitate hatching (19, 20). One interesting feature of male pregnancy is that the egg chorion breaks down relatively early in gestation, as paternal tissue proliferates and surrounds the embryos. Consequently, there may have been a selective pressure for the male to produce paternally derived hatching enzymes to facilitate choriolysis. If patristacin does function as a hatching enzyme, then it would represent an interesting case of convergent evolution within a gene family. Future functional assays of patristacin will be required to resolve this hypothesis.

In summary, our study has provided strong evidence that (*i*) patristacin serves an as-yet-unidentified role in the brood pouch of pregnant pipefish, (*ii*) cooption of patristacin for male pregnancy did not require gene duplication, and (*iii*) at least in the Syngnathidae, positive selection has driven amino acid changes potentially related to enzyme efficiency without loss of the ancestral function of the protein. Regardless, patristacin has been coopted to perform its function in a novel structure during the diversification of pipefishes and seahorses. Hence, the evolution of patristacin in male pregnancy provides an interesting case in which gene duplication was not necessary for gene cooption during the genesis of an evolutionary innovation, a finding of significance to our understanding of novel trait evolution.

Materials and Methods

Gene Identification and Sequencing. Total RNA was isolated from the brood pouches of pregnant ($n = 4$) and nonpregnant ($n = 4$) males of the Gulf pipefish (*S. scovelli*). Subtractive suppression hybridization (28) with a secondary screening procedure (29) identified 190 mRNA transcripts unique to the brood pouches of pregnant males, of which 95 were cloned and sequenced with standard methods. Of these sequences, 81% ($n = 77$) were identified via National Center for Biotechnology Information database BLASTx searches as members of the astacin subfamily of metalloproteases. A consensus sequence of aligned astacin-like amplicons was used to design pipefish-specific primers (Table 3, which is published as supporting information on the PNAS web site) that were used to obtain the complete mRNA sequence via RACE. Standard quantitative real-time PCR techniques (primers in Table 3) were used to confirm the differential expression of patristacin in replicate pregnant male *S. scovelli* ($n = 6$) compared with nonpregnant males ($n = 6$) and embryos ($n = 2$). Each quantitative real-time PCR was performed with samples in triplicate under the following cycling conditions: one cycle, 50°C, 10 min; one cycle, 95°C, 2 min; 40 cycles, 95°C, 15 sec, 60°C, 30 sec). A standard curve, derived from amplification of 18S in triplicate, was used to standardize the expression levels of the target gene. Details of PCR experimental conditions are available from the authors upon request. Reverse-transcribed, tissue-specific total RNA or genomic DNA was used as template for amplification of patristacin homologs in other teleost species (Table 1) with primers designed from the complete *S. scovelli* mRNA sequence (Table 3). Similar methods were used to amplify nephrosin homologues from syngnathid liver and kidney mRNA with primers designed from complete mRNA sequences of nephrosin from *D. rerio* and *Cyprinus carpio* (Tables 1 and 3). Astacin mRNA sequences from other taxa were downloaded from the GenBank database (Table 1), translated, and aligned to our nephrosin and patristacin consensus sequences with T-Coffee (30).

Evolutionary and Statistical Analyses. The evolutionary relationships among astacin genes were reconstructed in PAUP*v.4b10 (31) under a maximum parsimony optimality criterion, with branch lengths subsequently optimized via maximum likelihood. This phylogeny with branch lengths was used in subsequent tests for positive selection. If patristacin had undergone selection for a new function after gene duplication or cooption, we would expect to see an acceleration in the rate of evolution after these events. Furthermore, changes in the function of a gene are often (but not always) facilitated by alteration in the amino acid sequence of the protein via positive selection. We assessed evidence for variation in selection pressures along patristacin lineages and amino acid sites with the codon-based maximum-likelihood procedure implemented in the Codeml program in the PAML package (32) (http://abacus.gene.ucl.ac.uk/software/paml.html). Details of this program and associated tests for selection can be found in the online documentation of the program or in the original article describing the methods (32). We first tested for changes in selection pressure immediately after the gene duplication event that led to the appearance of patristacin in teleosts (lineage α in Fig. 3) and the cooption of patristacin for male pregnancy in Syngnathidae (lineage β in Fig. 3). For this test, we compared the likelihood fit of two sets of evolutionary models, one with ω constant among all lineages, the other with different ω values along the lineage(s) of interest (Table 2). With Codeml we then tested for variation in selection pressure among sites by comparing the likelihood fit of a series of nested models: one ratio (M0), neutral (M1), positive selection (M2), discrete (M3), β (M7), and β plus ω (M8) (Table 4, which is published as supporting information on the PNAS web site). Log likelihood values were compared with a likelihood ratio test, with

significance determined from critical values of the χ^2 distribution with estimated degrees of freedom as described (32). If the model with positive selection had a significantly better fit to the data, the posterior probability for positive selection at each site was estimated (33). A posterior probability ≥90% was considered moderate support and a posterior probability ≥95% was considered strong support that a site is under positive selection. In addition, we tested for positive selection among sites with the random effects likelihood (REL) method (34) via the Datamonkey (35) adaptive evolution server (www.datamonkey.org). The REL method is a modified parsimony-based method, which allows for rate heterogeneity in both dN and dS (36). REL analysis calculates two Bayes factors (BF), one for negative selection (dN < dS) and one for positive selection (dN > dS). We considered a BF ≥50 as moderate evidence and a BF ≥100 as strong evidence that a site is under positive selection. Whether maximum likelihood or parsimony-based methods are better for detecting sites under positive selection is a subject of debate (37, 38). To circumvent this controversy, we usually considered sites to be under positive selection only if (*i*) they met the initial BF and posterior probability limits and (*ii*) were identified in both REL and PAML analyses. Ten of the sites identified with REL ($n = 12$) and PAML ($n = 14$) methods met these criteria. We also identified two sites from the REL analysis that exhibited very strong evidence for positive selection (BF ≫ 100) and hence included them in the final count of 12 positively selected sites.

Protein 3D Model. We modeled the 3D structure of patristacin via the Swiss-Model homology modeling server (39) (http://swissmodel.expasy.org) with the fully resolved crystal structure of the *Astacus astacus* digestive enzyme (astacin) as a template. Amino acid substitutions in patristacin were compared with aligned substrate-binding and conserved structural residues in astacin. These sites and those under positive selection were visualized and annotated in Cn3D v. 4.1 (40).

Molecular Kits. TOPO TA (cloning), GeneRacer (RACE), SuperScript III (reverse transcription), and TRIzol (RNA isolation) were from Invitrogen (Carlsbad, CA). PCR-Select (subtraction suppression hybridization) was from Clontech (Mountain View, CA).

We acknowledge K. B. Mobley, C. Partridge, Z. Cress, and C. Small for sample collection (pipefishes and sticklebacks). S. Coleman, R. Honeycutt, T. Ward, K. B. Mobley, and two anonymous reviewers provided helpful comments on previous drafts. Funding was provided by a grant from the National Science Foundation (to A.G.J.).

1. Wagner GP, Lynch VJ (2005) *J Exp Zool B* 304:580–592.
2. Carcupino M, Baldacci A, Mazzini M, Franzoi P (2002) *J Fish Biol* 61:1465–1480.
3. Haresign TW, Shumway SE (1981) *Comp Biochem Physiol A Physiol* 69:603–604.
4. Watanabe S, Kaneko T, Watanabe Y (1999) *Cell Tissue Res* 295:141–149.
5. Quast WD, Howe NR (1980) *Comp Biochem Physiol A Physiol* 67:675–678.
6. Azarello MY (1991) *Bull Mar Sci* 49:741–747.
7. Wilson AB, Vincent A, Ahnesjo I, Meyer A (2001) *J Hered* 92:159–166.
8. True JR, Carroll SB (2002) *Annu Rev Cell Dev Biol* 18:53–80.
9. Ohno S (1970) *Evolution by Gene Duplication* (Springer, Berlin).
10. Lynch M, Katju V (2004) *Trends Genet* 20:544–549.
11. James LC, Tawfik DS (2003) *Trends Biochem Sci* 28:361–368.
12. de Souza FSJ, Bumaschny VF, Low MJ, Rubinstein M (2005) *Mol Biol Evol* 22:2417–2427.
13. Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait JH (1999) *Genetics* 151:1531–1545.
14. Lynch M, Conery JS (2000) *Science* 290:1151–1155.
15. Jensen RA (1976) *Annu Rev Microbiol* 30:409–425.
16. Copley SD (2003) *Curr Opin Chem Biol* 7:265–272.
17. Bond JS, Benyon RJ (1995) *Protein Sci* 4:1247–1261.
18. Titani K, Torff JJ, Hormel S, Kumar S, Walsh KA, Rodi J, Neurath H, Zwilling R (1987) *Biochemistry* 26:222–226.
19. Inohaya K, Yasumasu S, Ishimaru M, Ohyama A, Iuchi I, Yamagami K (1995) *Dev Biol* 171:374–385.
20. Yasumasu S, Iuchi I, Yamagami K (1988) *Zool Sci* 5:191–195.
21. Tabas JA, Zasloff M, Wasmuth JJ, Emanuel BS, Altherr MR, McPherson JD, Wonzney JM, Kaplan FS (1991) *Genomics* 9:283–289.
22. Ishmael FT, Norcum MT, Benkovic SJ, Bond JS (2001) *J Biol Chem* 276:23207–23211.
23. Kijimoto T, Watanabe M, Fujimura K, Nakazawa M, Murakami Y, Kuratani S, Kohara Y, Gojobori T, Okada N (2005) *Mol Biol Evol* 22:1649–1660.
24. Hung C-H, Huang H-R, Huang C-J, Huang F-L, Chang G-D (1997) *J Biol Chem* 272:13772–13778.
25. Bordo D, Argos P (1991) *J Mol Biol* 217:721–729.
26. Foskett JK, Scheffey C (1982) *Science* 215:164–166.
27. Varsamos S, Nebel C, Charmantier G (2005) *Comp Biochem Physiol A Physiol* 141:401–429.
28. Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) *Proc Natl Acad Sci USA* 93:6025–6030.
29. von Stein OD, Thies WG, Hofmann M (1997) *Nucleic Acids Res* 25:2598–2602.
30. Poirot O, Suhre K, Abergel C, O'Toole E, Notredame C (2004) *Nucleic Acids Res* 32:W37–W40.
31. Swofford DL (1999) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA).
32. Nielsen R, Yang ZH (1998) *Genetics* 148:929–936.
33. Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) *Genetics* 155:431–449.
34. Kosakovsky Pond SL, Frost SDW (2005) *Mol Biol Evol* 55:1208–1222.
35. Kosakovsky Pond SL, Frost SDW (2005) *Bioinformatics* 10:2531–2533.
36. Kosakovsky Pond SL, Muse SV (2005) *Mol Biol Evol* 22:2375–2385.
37. Suzuki Y, Nei M (2001) *Mol Biol Evol* 18:2179–2185.
38. Anisimova M, Bielawski JP, Yang Z (2002) *Mol Biol Evol* 19:950–958.
39. Schwede T, Kopp J, Guex N, Peitsch M (2003) *Nucleic Acids Res* 31:3381–3385.
40. Hogue C (1997) *Trends Biochem Sci* 22:314–316.
41. Kornienko ES (2001) *Russ J Mar Biol* 27:S15–S26.
42. Rawlings ND, Barrett AJ (1995) *Methods Enzymol* 248:183–228.